

CD-ROM 長野県植物誌資料集テスト版 (植物資料データベースの問題点) (金井弘夫)
Hiroo KANAI: Introducing the CD-ROM Flora of Nagano Prefecture, β -Version (Some Points in Regard to Flora Database Construction)

長野県植物誌は同誌編纂委員会 (清水建美委員長) により1997年に出版されたが, このたびその土台となった約42万件 (そのうち標本レコードは175,975件) のデータベースを収容した表記のROMが制作された (清水2002). このROMは「テスト版」の名のとおり, 貧乏集団のやりくりの末, 最小限の機能を持たすにとどめて, 内部利用のために少数部が作られたもので, 一般頒布の普及版へ向かってこれから取り組むことになる. しかしいくつかの新しい試みがなされており, 普及版と関連してフロラや資料のデータベースを作っておられる方々にもご理解を得る目的で紹介すると共に, 制作に当たって気付いた植物資料データベースの問題点について述べる.

本ROMは内蔵の解凍ソフトでユーザーの機器に移植する (Windows対応). 機能の概略は次のとおりである.

(1)レコードの検索とリスト表示. (2)任意の植物の分布図の表示. (3)学名の全綴り (著者名を含む) の表示. (4)外部データによる分布図の表示. 分布図は日本全域対応. 印刷はどの機能についても可能である.

(1)検索は植物名, 地区, 産地, 位置, 高度, 年代, 採集者など, 23項目 (Table 1 参照) のすべてについてインデックス検索が可能である. 複数項目にわたるインデックス検索はできないが, 検索対象以外の項目については, 複数項目にわたる逐次.and. 検索が行える. また年代については, 期間を指定した検索が行えるので, 分布状態の変遷をたどることが出来る (Figs. 2-4).

リストの縦横スクロールはできないが, 横方向については表示項目, 表示順序, 表示幅を任意に指定できるので, 必要項目を画面内に納めることは容易である. また単票表示によってレコードの全内容を一覧できる. 縦方向についてはカーソル行先頭, 頁単位ジャンプ等により, 任意の表示状態にできる. 印刷は画面外のデータも可能である.

(2)分布図は, リスト画面で表示されている植物について, 直ちに表示される (Fig. 1). 分布マークの形, サイズ, 色, 優先順位は選択可能である. 地図は海岸線, 都府県界, 等高線 (200, 1500, 2500 m) が選択表示でき, 任意の位置にメモを記すことができる. 垂直分布図 (東西, 南北) も同時に作図され, その表示の可否と表示位置を選択できる. 縮尺を示すスケールバーは, 地図の縮尺に応じて適当なものが自動的に選択表示される. 地図上の分布点をクリックすれば, その根拠となったレコードが表示される. 位置座標のつけ誤りなどのため, 不都合な位置に分布点が現れる場合には, これを一時的に消去する機能をもつ. この機能はリストについても使える.

地図は北方四島を含む日本全域に対応し, 県別に用意された枠を指定した上, 任意の大きさや縮尺に作り直すことにより, 日本のどこでも分布図を描くことができる. 経緯度線または基準メッシュ (二次メッシュ) を選択表示することが可能である. 長野県については, 500 m 等高線と市町村界を表示することができる.

最近の地図ソフトでは, 地勢図, 地形図, 住宅地図などをオーバーレイできるものが出回っているが, これらは詳細すぎて発表用原図としてはかえって思わしくない場合がある. 本ROMの地図は簡単な線画のみで構成されており, プリント時にはカラー画面を自動的に白黒画面に変換して出力する機能を持っているので, 印刷発表用原図として使い易いと考えている. 分布図の印刷は, 引き続く植物名については連続して出力できる.

(3)長野県植物誌データベースは1970年代から蓄積が始まっており, わが国最古の起源を持っているため, 項目ごとの文字数制限が厳しかった. とくに和名や学名については, 綴りのごく一部しか収容できない (Table 1). これを補うため, 和名や学名の全綴りを収容する「和名学名対照表」を内蔵しており (Table 2), 植物誌リストと常に連携させて

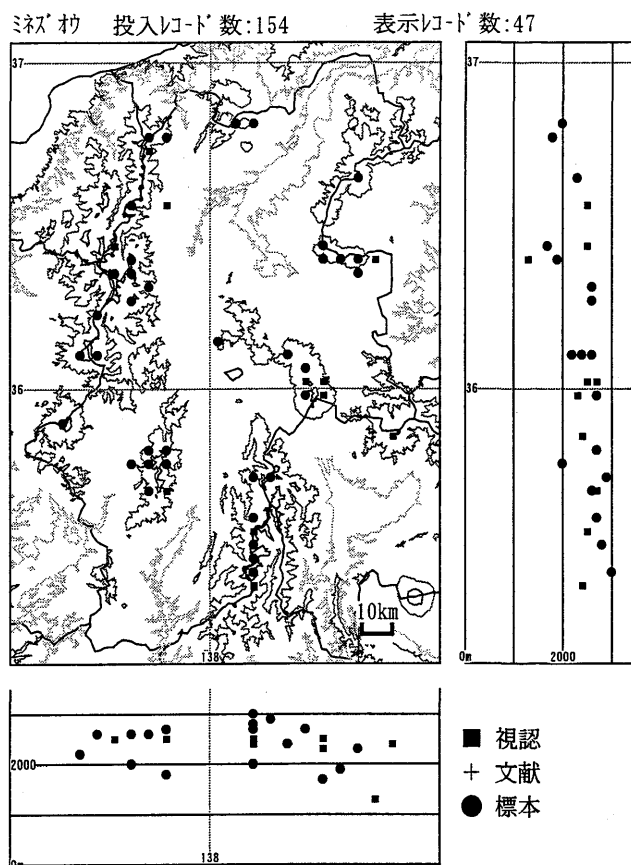


Fig. 1. ミネズオウの全レコード. 等高線200, 500 (これは長野県のみ表示可能), 1500, 2500 m. 経緯度線1度間隔. 記録メッシュと表示メッシュは Locality Index. 垂直図3000 m用.
Distribution map of *Loiseleuria procumbens*. Contours 200, 500, 1500 and 2500 m. Recording and display mesh are Locality Index. Vertical map for 3000 m.

ある。リスト画面からの指示により、この対照表から和名や学名（著者名を含む）の全綴りが表示され、異名もある程度わかるようにしてある。またリストを印刷するとき「フロラリスト印刷」を指定すれば、植物誌データからではなく、対照表から和名学名の全綴りを出力できる。

和名学名対照表は、長野県植物誌および環境庁植物目録（1994）等から取り込んだ。将来は日本植物全体を対象として更に他の文献からも補う予定で、著者名や出典の検索にも利用できる仕様にしてあるが、このROMで

はその一部を使っているに過ぎず、内容はまだ不十分である。

(4) このROMのもう一つの特徴は、長野県のみならず、日本のどの地域についても、ユーザー自作のデータを使って分布図（垂直図を含む）を作れることである。このためには「見出し名（植物名）」「座標系記号」「位置座標」「高度」「資料記号」のテキストデータを用意すればよい。最近の植物誌では地域分布図の表示はもうあたり前であるが、それらは互換性がない。このROMを利用すれば、共通の基図の上に分布図を描くことができ、

隣り合う地域の分布図を連結描画できる。この目的のために、位置座標としては **Locality Index**, 経緯度 (分), 三次メッシュ, 五倍メッシュのいずれについても受入れ可能であり, しかもこれらを混用できるようにしてある。したがって, 植物のみならず自然現象, 文化現象などの分布表示に活用できる (Figs. 5-7)。たとえば危惧種の選定資料を得る場合, 五倍メッシュ程度の表示を使って, その一定割合以下の出現率を持つ種類を候補とするような作業を, 容易に行うことができた。サンプルとして 金井の普通植物分布調査の全国データの一部が, 使えるようにしてある。

冒頭に述べたように, 本 CD-ROM は関係者の内部利用を目的として, 少数部作られた非売品である。近い将来データの訂正追加を行った上, 一般頒布用の普及版が計画されている。その際には, 長野県のみでなく, 日本全域の植物分布データを, 可能な限り取り込む予定である。制作費節約のため, マニュアルは委員会側で作成したので, 市販のものにくらべて一風変わったものになっているが, ソフトやデータの内容については, これを読めば理解の助けになるだろう。本 ROM には絶滅危惧種などの公表をはばかるデータが含まれており, 一般頒布の際にはしかるべき処理をほどこす予定である。それ迄は本 ROM の機能や CD によるフロラ出版に関心ある方, あるいは自分の持つデータを使ってみたい方は, 下記に問い合わせられたい。

長野県植物誌資料集作成委員会, 390-0312 松本市岡田松岡 211-3 清水建美, または 184-0013 小金井市前原町 5-8-7 金井弘夫

Figs. 1-7 に, いろいろな機能による出力例を示す。

以下, CD-ROM の制作中に出会った植物データベースのいろいろな問題点について記す。

○一般頒布用 CD-ROM の構想

本 CD はすでに日本のどの地域にも対応できるようになっているので, 一般頒布用 CD は日本のどの地域の植物データも取り込むつもりであり, ゆくゆくは長野県以外の地域の植物誌データも取り込んで, 日本植物誌資料集という形に発展させることを指向している。

そのためには, 地域植物誌データベースや広域標本室データベースの参加が不可欠であるが, それらをそのまま取り込めば済むというものではない。本 CD の仕様に変換および書き換えを行った上でないと, 有効利用はできない。このことは統一仕様を要求するものではなく, 目的の異なるデータベースが共存するためのものである。以下に問題点を記す。

○記録用データベースと検索用データベース
一般頒布用 CD の参考とするため, いくつかの機関のデータベース作成事例を教えてください。標本や資料の記録には「元の記述に忠実に」ということが前提とされる。ところが「元の記述」は記録者によって (場合によっては同一人でも) 多種多様で, これを忠実にデータベースに作れば, 検索という目的には不適當である。たとえ項目の仕分けをきめ細かく行っても, その中身の多様性は失われることがない。とくに最近の表ソフトでは, 項目長の制限を意識しないで済むので, 検索目的のために決まった位置に目的の単語をはめ込む努力はなされない。長野県の場合には初期の仕様を長らく維持してきたので, 項目ごとの文字数制限が厳しく, 元の記述から逸脱しても, 検索目的に合った単語をそこへはめ込むようにして来た。その結果, 「オリジナルな資料と合わない」という不満をぶつけられたこともある。しかしながら今回の CD-ROM のような, 項目ごとのインデックス検索を目的としたデータベースでは, こういうやり方でないと目的を達しないということを痛感した。一般頒布用 CD-ROM は, インデックス検索を目的とするものなので, この方針を維持するつもりである。したがって地域フロラデータベースをそのまま取り込むのではなく, 本 CD の仕様に合わせた項目分割とデータの書き換えを行なうことになる。ただ, 現在の仕様は窮屈すぎるので, よそからの参加をし易くする為にも, 項目幅の拡張 (とくに数字をデータとする項目) や項目の新設 (「日」「行政コード」など) を予定している。

○位置座標

本 CD の特徴の一つは分布図作図にあるが, そのためには産地の位置座標が必要である。長野県植物誌では **Locality Index** (5 倍メッシュ相当) を用いている。当初はメッシュによる

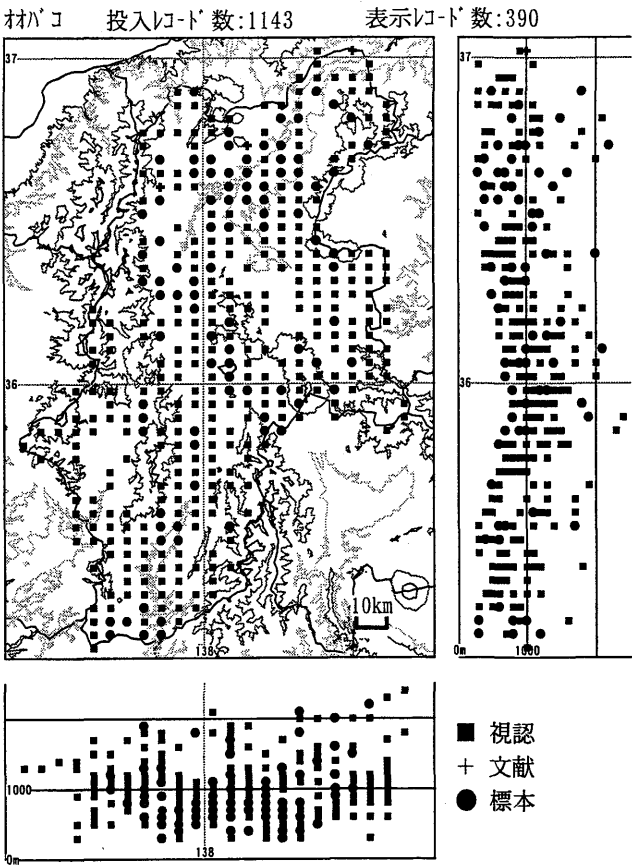


Fig. 2. オオバコ的全レコード. 等高線200, 500, 1500, 2500 m. 経緯度線1度間隔. 記録メッシュと表示メッシュは Locality Index. 垂直図2000 m 用.
Distribution map of *Plantago asiatica* (total record). Contours 200, 500, 1500 and 2500 m. Recording and display mesh are Locality Index. Vertical map for 2000 m.

Table 1. 長野県植物誌データベースの項目 Structure of flora database

項目名	桁数	内容	項目名	桁数	内容
和名	10	片仮名	年	4	西暦
地名	12	なるべく小範囲	月	1	数字, 英字 (X, Y, Z)
地名読み	9	片仮名	状態	1	0 蕾, 1 花, 2 果実, 3 葉
系	1	メッシュ系記号	資料	1	7 標本, 8 文献, 9 視認
位置座標	8	数字	標本室記号	4	英字
地区	3	英字の記号	標本番号	6	下6桁
市町村	8	漢字	科順位	6	分類系内の序列
高度	4	m 単位	科名	10	英字
氏名	12	漢字	属	10	英字
氏名英字	7	姓を先頭	種小名	6	英字
採集番号	4	下4桁	最下小名	6	種内群の最後の小名
			確認記号	1	標本確認の印

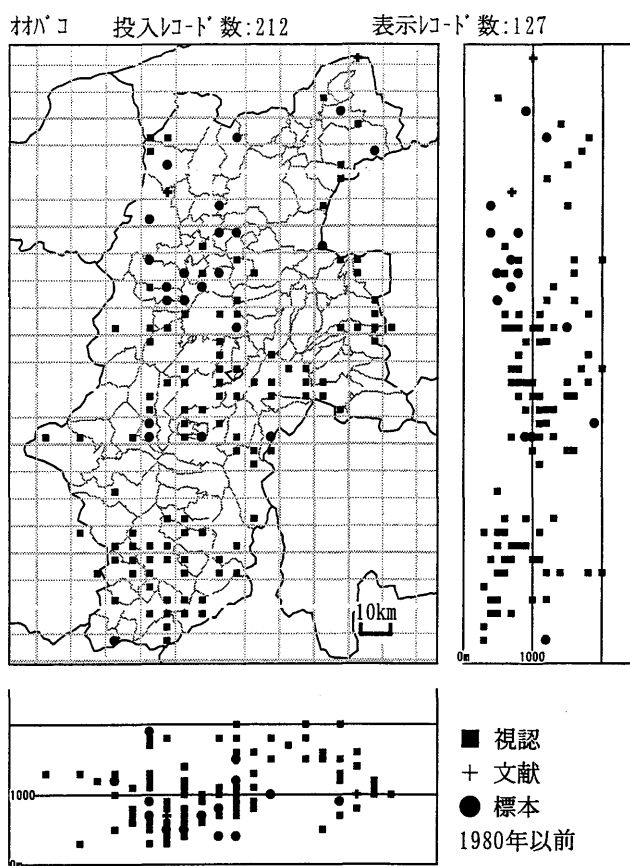


Fig. 3. 年代検索の例. オオバコ. 1980年以前. 市町村界 (これは長野県のみ表示可能). 網目は二次メッシュ. 記録メッシュと表示メッシュは Locality Index. 垂直図 2000 m 用.
Distribution map of *Plantago asiatica* before 1980.

Table 2. 和名学名対照表の項目 Structure of plant name thesaurus

項目名	桁数	内容	項目名	桁数	内容
採用和名	1	植物誌採用の和名の印	雑種	1	雑種記号
和名	10	植物誌の和名	小名全綴	68	全小名(ランク名含む)
和名全綴	18	和名の全綴り	最下小名	6	最後の小名
和名清音	18	50音配列用字列	著者名	52	最後の小名の著者 (中間の小名の著者は省略)
科順位	6	分類系内の科の序列	出版年	4	学名の出版年
科名	10	科の学名	由来	2	外来, 栽培など
採用学名	1	植物誌採用の学名の印	備考	10	メモ用
雑種属	1	雑種属の印	作業用	2	作業用マーク欄
属全綴	20	属の学名			

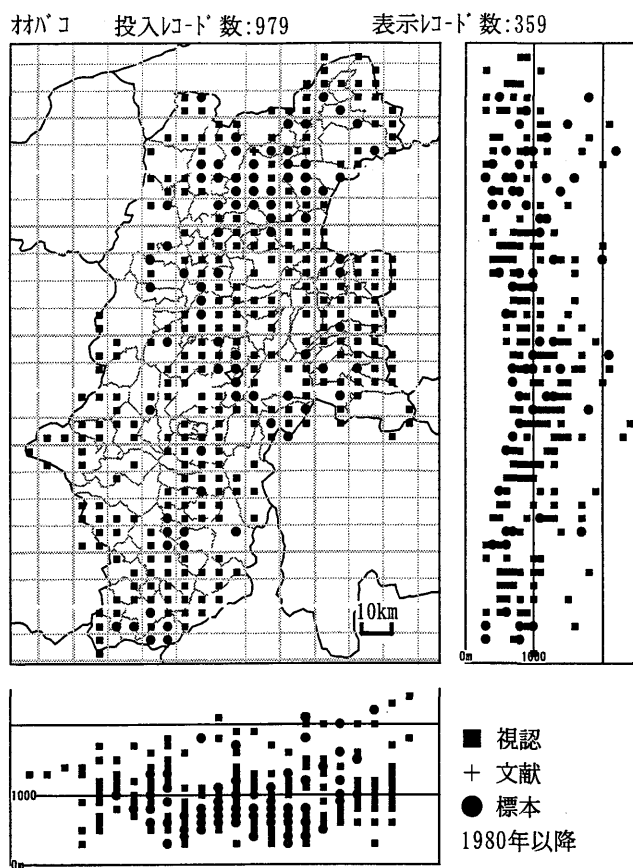


Fig. 4. 年代検索の例. オオバコ. 1980年以降. 市町村界. 網目は二次メッシュ. 記録メッシュと表示メッシュは Locality Index. 垂直図2000 m 用.

Distribution map of *Plantago asiatica* after 1980.

位置記録に慣れないせいもあって、「こんなに細かいメッシュではやって行けない」という声もあった。しかし最近では環境省の調査の影響で、三次メッシュ（いわゆる1キロメッシュ）のような精細なメッシュによる記録が増加している。それ自体は結構なことであるが、これですべての資料の位置座標が表示できるとは思われない。なぜならばこのような「精密な」位置決めは、採集者自身がその場で記録するからこそできるのであって、これ迄蓄積された他人の資料の産地名からでは不可能だからである。地名にはさまざまな広さがあり、採集地点の記録にはそこから一番近いところの地名、あるいは自分になじみのあ

る地名を引用するだけなのだから、地名の表記されている三次メッシュが採集地点だとするわけにはゆくまい。地名から位置座標を得るには、もっと疎なメッシュ系を用いるのが妥当である。その最も一般的なものは経緯度の分の単位だろう。分メッシュのサイズは中部日本では東西約1.5 km 南北約1.8 kmである。本当はこれより疎な5倍メッシュ (Locality Index はこれに相当し、東西約5.7 km 南北約4.6 km) 程度が適当と思う記録の方が多いのだが、このメッシュ系は今やなじみが薄くなっている。たとえば「～村」とだけしかない産地記録が古い資料や文献には少なからずあり、これを無視すると時間的変遷

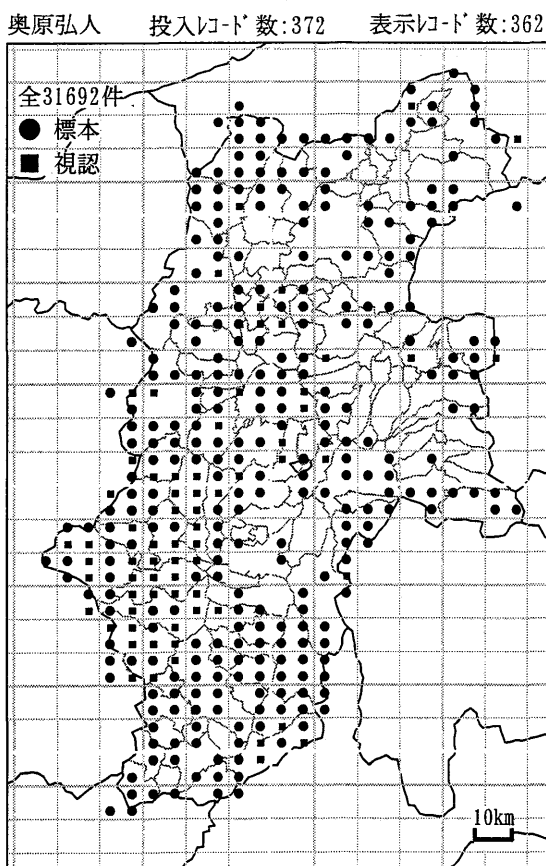


Fig. 5. 外部データ利用の例. 奥原弘人氏の全資料の分布. 市町村界. 網目は二次メッシュ. 記録メッシュと表示メッシュは Locality Index. 植物誌データベースを氏名でインデクス付けし, 氏名項目が「奥原」のレコードを「外部データに保存」により, 外部データ利用の仕様に合わせて抽出 (31,692件). これを表ソフトで読んで植物名項目を「奥原弘人」に書き換えた上, 位置座標にユニークインデックスをかけ, テキストファイルとした上, 外部データ利用機能で分布図を得たもの.

Example of import data utility. Map of Mr. Hiroto OKUHARA's activity.

をたどれない心配があるが, 分メッシュで位置座標を与える気にはとてもなれないのである. どれを使うにせよ, 位置座標系を一つに限定すれば, 記録の精度がわからなくなってしまう. だから本 CD では系の異なる位置座標も容認し, しかも混在してもよいことにしている. そのかわり, どの座標系による記録なのかを, 位置座標系記号で表示しておく. 分布図作図の際, あらかじめ表示メッシュ系

を指示しておけば, ソフトは記録されたメッシュ系記号によって経緯度に換算した値を, 指示された表示メッシュ系の値にもう一度換算して分布点マークを表示するようにしてある. このとき, 表示メッシュより疎な記録メッシュ系のデータも取り込むか排除するかを, 作図の目的に応じて選択できるようにしてある. 産地名に対する位置座標を付加する作業は, 本 CD を利用する上で最も重要でかつ手

間のかかる作業だろう。これは「地図で見る日本地名索引CD-ROM」(ABOC社1998)を利用すれば、かなり容易になる。

○地域名(県名と市町村名)

旧国名は「忠実な記録」としてはもちろん大切だが、「検索用」としては役に立たないと言わざるを得ず、県名で代替するほかはない。県内では、郡名は地名の特定にあまり有用ではないので、省いても差し支えなからうというわけで、市町村名だけを項目として選んである。ただし「市」は範囲が広すぎて、位置特定には不便である。市内の町名や区名の方が有用だが、これは同じ県内で他に同じ名前がある場合が少なくない。結局、市名は使わざるを得ず、その中の町名や区名は、必要なら地名項目に書き込むことになる。

近い将来、町村合併が推進されて市がたくさんできるようになると、市町村名の項目に何を収容するかは考え直す必要が出てくるだろう。時代の異なる記録を収容するのだから、その資料の作成年代の市町村名が使われていて、これを記録せざるを得ず、全体としては不揃いになる。これを「現在」に合わせて一々書き換えるということは、際限のない作業であり、賢明とは思われない。別に「行政コード」の項目を設けておいて、県以下の自治体の番号を記録しておく方がよい。この番号も自治体統廃合に伴って変わるが、こちらは常に「現在」に合わせて書き換えて行き、市町村名は元のままにしておけばよい。要するに市町村名は元のままを保存し、行政コードで現状に対応するというやり方がよさそうに思う。

○産地名

位置特定のためには、範囲の最も小さな地名(小字名)や地物名を選ぶ必要がある。しかしこれではわかりにくいことがあるので、上位の地名(大字名など)も入れておいた方がよい。なお「大字」とか「字」とかいう文字は、データとしては区切り記号として機能するにすぎないので、省いてよいと思う。

この項目では、「元の記述に忠実」であることと「検索を目的」としたデータが最も矛盾する。一番簡単な例を挙げると、(1)「槍ヶ岳」(2)「槍ヶ岳」(3)「槍ヶ岳」(4)「槍が岳」のような場合である。これらの綴りは記録者

の意見なり習性なりを反映するものなので保存したいのだが、このままでは検索でもソートでも同じところに集まってくれない。したがって「統一した」記述に書き換える必要がある。そこでどの綴りに「統一する」かとなると、衆目の一致するところは(1)「槍ヶ岳」である。けれども検索する立場に立つと、「ヶ」という文字はたいへん使いにくい。この文字は全角の小文字で、通常は漢字辞書を介して呼び出すことになり、検索に先立って利用者がそういうことに気づくことは少ないだろう。おまけにリストのためのソートを行うと、全角片仮名の「ン」より後に出現してしまい、きわめて不自然である。だから私は、衆議に逆らっても(2)「槍ヶ岳」に「統一」してしまう。

もう一つ例を挙げると、「A~B~C」というように、区間を示す産地記述である。これはきわめて頻繁に出会うものだが、資料一点の産地記録としてはあり得ない筈である。高度記録でも同様である。分布図を目的とするのだから、「点」としての位置座標を与える必要がある。私はこの内どれか一つ、多くの場合は先頭、の地名の位置座標を与えることにしている。前記の位置座標の見出しの下で、「精密さ」について懐疑的なのは、このような「いい加減な」記述に対して「精密な」位置座標を与えても、自己満足に過ぎないからである。自分の資料に対してはいくらでも「精密な」座標を付加できるが、分布を扱うときにはこういう「いい加減な」記録の方がはるかに多いのだから、精密な座標の方を良しとするのは考えものである。表示メッシュは三次メッシュ、分、Locality Index(五倍メッシュ相当)のどれかを選択できるが、「精密」だからと三次メッシュを表示に使えば、Locality Indexのような疎なメッシュ記録を除外するか否かを選択する必要がある。さもないと分布点の出方が不自然になる。

○学名

長野県植物誌データベースの植物名の文字項目幅はきわめて狭い(Table 1)。しかしこれらを組み合わせれば、ほとんどの場合植物名を特定し得るし、植物名を科の下で学名順に配列することは可能である。最下小名とは、種内分類群の最後のランクの小名のことであ

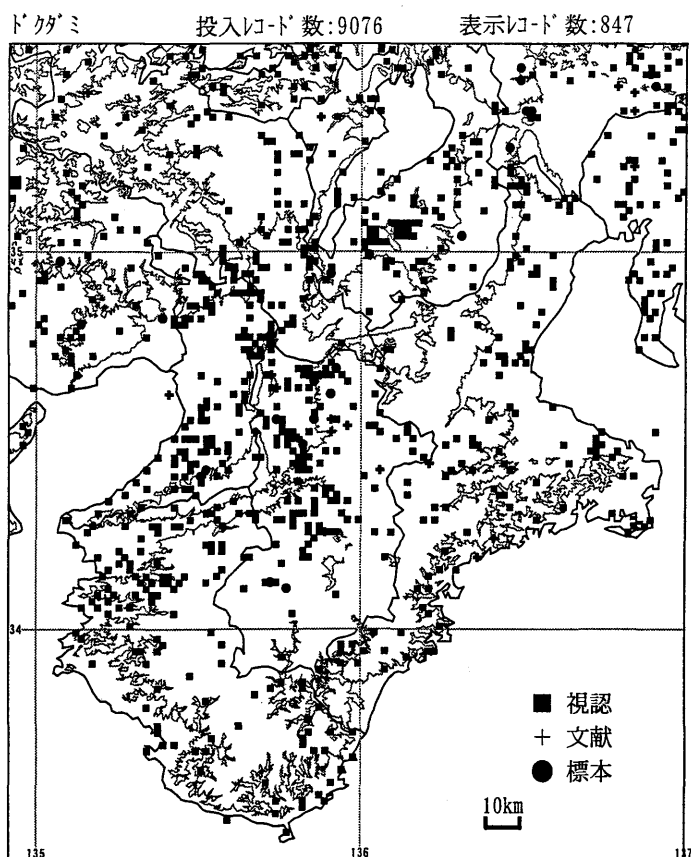


Fig. 6. 外部データ利用の例. 近畿地方のドクダミの分布図. 金井の普通植物分布調査の結果. 等高線 200, 1500 m. 経緯度線 1 度間隔. 記録メッシュは Locality Index, 分, 三次メッシュの混合. 表示メッシュは分. Example of import data utility. Distribution map of *Houttuynia cordata* in Kinki district, central Japan.

る. 科順位とは, フロラリスト出力の際, 科を分類系内の順序に配列するための順位表で, 長野県の場合には Cronquist 系の配列番号が入っている. 他の系を用いる場合には, ここにそれに応じた数字を入れ換えればよい. 色々な系による序列番号は別に用意されているが, これは本 CD には内蔵されていない. 普及版では考慮する必要がある.

○和名学名対照表

前記のとおり, データベースの植物名項目は, どれについても全綴りを収容するには足りない. われわれはほとんどの場合和名で用が足りるし, このままでも検索や分類順配列

には差し支えない. しかしリストとして表に出すのは気が引ける. そこで植物名の全綴りを収容できるデータベースを別個に用意した (Table 2). この対照表と植物誌のデータは常に連携しており, リスト画面から対照表を呼び出せば 和名項目を介して学名の全綴りが表示され, 異名関係もある程度わかるようにしてある. 和名と学名は 1:1 に対応する必要があるので, 本 CD で採用する和名と学名のセットにはマークがつけられている. データベースがある程度たまった段階で, 和名をこの対照表にマッチングさせ, アンマッチのレコードについては採用和名に書き換えたり,

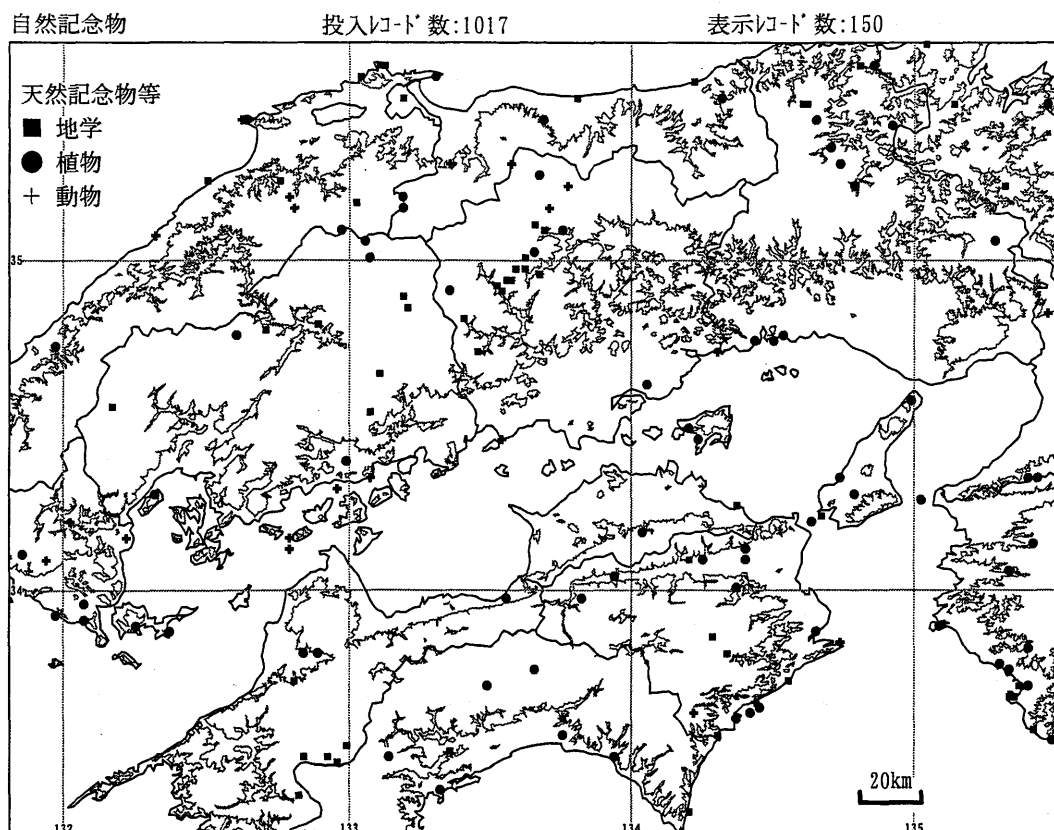


Fig. 7. 外部データ利用の例. 中国・四国地方の天然記念物等. GPS 全日本ロードマップ (徳間書店 1991) 等から採録. 等高線 200, 1500 m. 経緯度線 1 度間隔. 記録メッシュと表示メッシュは分. Example of import data utility. Distribution map of natural monuments in Chugoku and Shikoku districts, western Japan.

対照表に植物名を付け加えたりする操作をその都度行っておいた. このとき最も厄介なのは和名の同名異物で, マッチングをかけてしまえば本来はどちらだったかわからなくなるので, そういう和名についてはアンマッチになるような細工をしておかねばならない. けれどもあらかじめそういう和名を見つけるのは, 大変むずかしい. 和名にはルールがないから方言名などを使われたら混乱の元になる.

リストを印刷出力する際, 「フロラリスト印刷」の機能を指定すれば, 植物名についてはこの対照表に基づいて和名, 学名 (著者名を含む) の全綴りを出力し, しかも同じ和名なら最初の一件のみが表示されるようにして

ある. 科名についても同様である. したがって出力項目を適当に選んで, いわゆる植物目録形式の印刷出力を行うことが可能である.

植物標本資料データベースを作っている多くの機関では, すでにこういうシソーラスを用意していて, 和名を入力すればその学名が自動付加されるような仕組みになっているところが多い. その上, 異名や著者名, 外来品などの情報も, レコードごとに付け加えられる. これは優れた工夫だと思うのだが, 一方ではレコード一件ずつの情報量が膨らんで, いわゆる「重たい」データベースになるおそれがある. 私は, 資料固有の情報と, それが属する分類群に関する情報を分離する方が, 資料データベースの扱いが楽ではないかと考

